It's starting to shape up what tiny corp's product will be. It's not much of a change from what we sell and do now, but the vision is clearer.

Every month, we see these LLMs become more and more human. However, there's a major difference. They do not learn. Everyone has the same Claude/Codex/Kimi, with the same weights, the same desires, and the same biases. If current trends continue, the collapse in diversity will be staggering. To paraphrase:

*I think there is a world market for maybe five people.*

This is not the future I want to live in.

---

If trends continue where there's a single model with frozen weights and all learning is in-context, the cloud will win. Except in some highly latency sensitive (fighting robots) or connectivity critical (self driving cars) environments, it will be cheaper to run in batch on the cloud.

The enshittification that came to the web won't be the driving force to local models. We either live in a world where open models are so bad even user-hostile closed models are better, or open models are good enough, and competition to run them through sites like openrouter will prevent enshittification.

The only way local models win is if there's some value in full on learning per user or organization. At that point, with entirely different compute needing to run per user, local will beat out cloud.

The open question is if everything that's unique about you can fit in a 10 kB CLAUDE.md. If that's true, we have a pretty sad future ahead. It's the Attack of the Clones, swarms of identical minds you have no say over all varying in a small boxed-in way. This isn't learning, it's *costuming*. Everyone who has used these things knows how little of an impact prompting makes compared to the model. It's the Internet funneled into a little box you can edit on your profile. Write 3 paragraphs about what makes you unique.

We have to build for a future where that isn't true. 90% of people will choose the cloud, and what they will find is that they are no longer meaningfully in the loop. The dream is an AI product that will do your job for you while you continue to get paid. But this cannot exist, that's way too much of a fee to pay to the middleman. If you choose the homogenous mind, you are superfluous and will be cut out. Is there anything uniquely valuable about you? And I mean honestly, not the self-esteem pumping speeches you may have heard in school. If there's not, I have some bad news for you…

---

We already sell the hardware. Consumer GPUs still are the cheapest way to run models. There's tons of work required on the infrastructure. The frontend will be the future iterations of OpenClaw and opencode. But the key distinction from what you have today is that your tinybox will learn. It will update the weights based on its interactions with you. Like living things.

This is many years away. Currently, we are focused on large LLM training (even running these things is hard, have you tried to use vLLM not on NVIDIA?) and generic infrastructure for driving GPUs. But this is the long term idea.